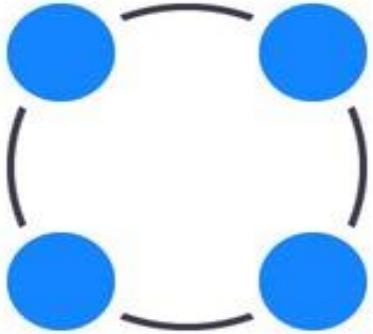


Реконструкция последовательности белка с учетом вариаций

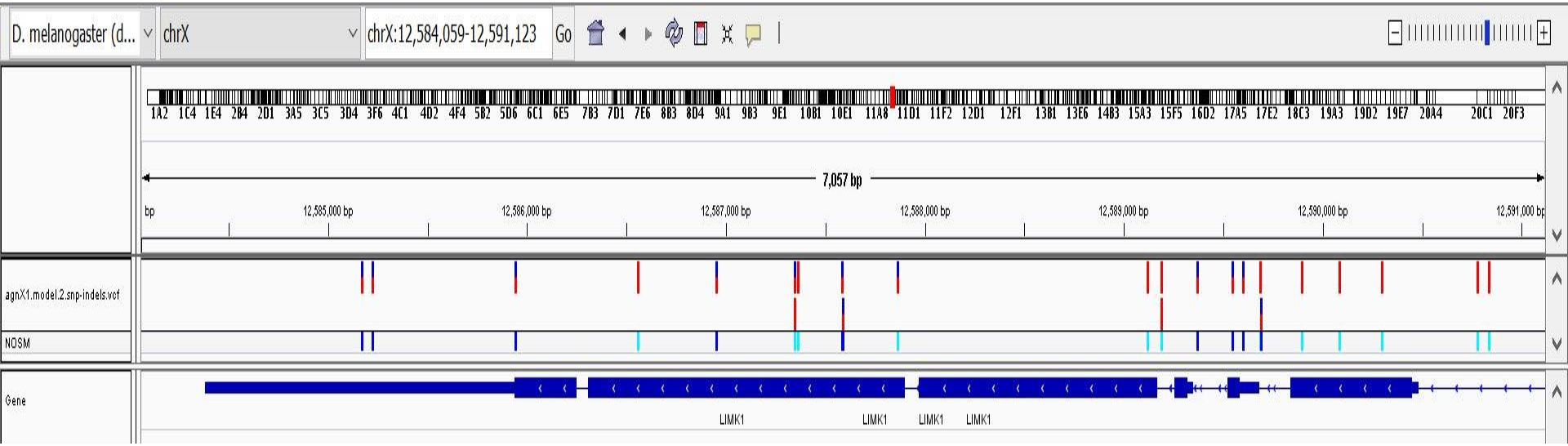
Научный руководитель: Захаров Геннадий,

Участники проекта: Нестеренко Максим, Цуринов Петр, Кожевникова Ольга, Коляденко Илья

Научные консультанты: Зуева Мария, Альперович Михаил

Цель работы

Создать инструмент для восстановления аминокислотной последовательности белка, с учетом всех мутаций, локализованных в белок-кодирующих участках.



Задачи

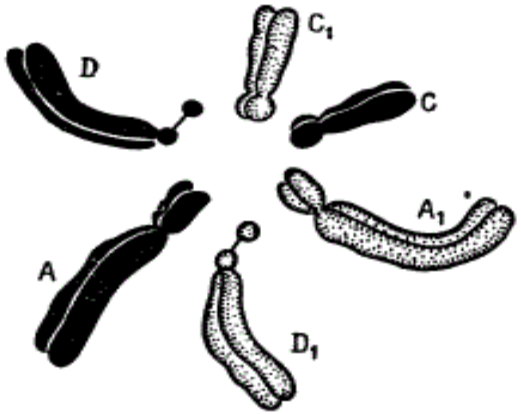
- Изучение форматов файлов NGS, используемых в эксперименте (FASTA, VCF, GTF)
- Применение мутаций на референсную последовательность
- Восстановление аминокислотной последовательности с учетом мутаций

Различные варианты локализации мутаций

1/1

0/1

1|0



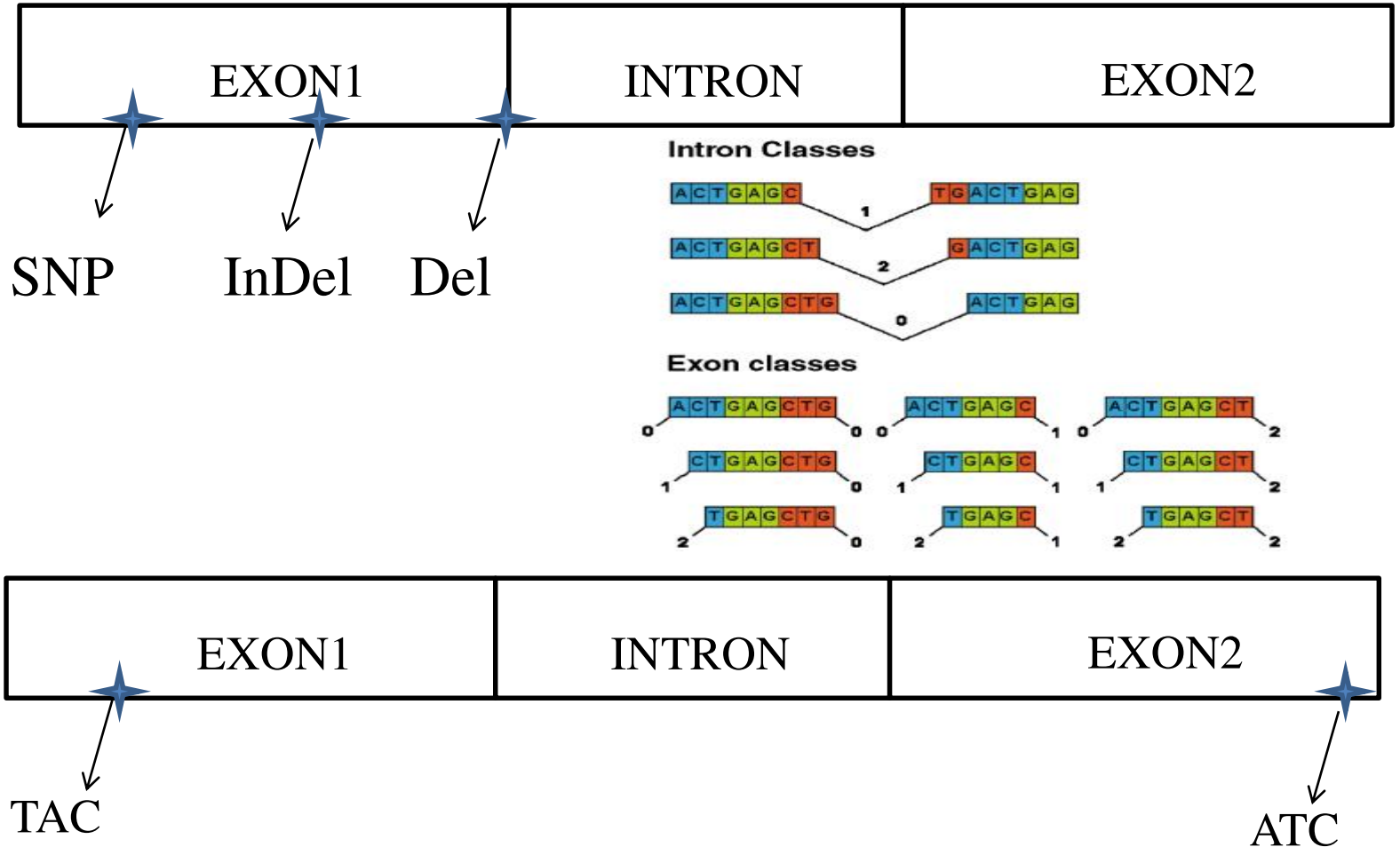
1|0

Каждая гетерозиготная нефазированная мутация приводит к **удвоению** возможных вариантов последовательностей

Входные данные

- Референсная последовательность дрозофилы (Fasta)
- Аннотации генов (GTF)
- Список мутаций (VCF) – 30 мутаций, из них:
  - 15 гетерозиготных, нефазированных, из них:
  - 10 в экзонах, из них:
  - 6 в белок-кодирующей части
- Таким образом, гипотетическое пространство вариантов равно $2^6 = 64$

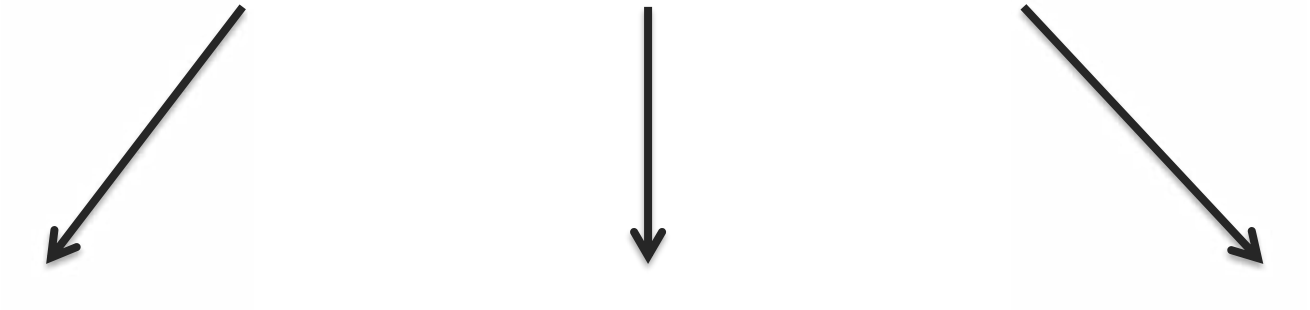
Изменение интрон/экзонной картины



Подводные камни

- Влияние мутаций друг на друга
- Изменение положений экзонов и стартового кодона при мутациях
- Направление считывания

Языки, используемые в работе:



Python

- Быстрый прототип
- Библиотека biopython

Java

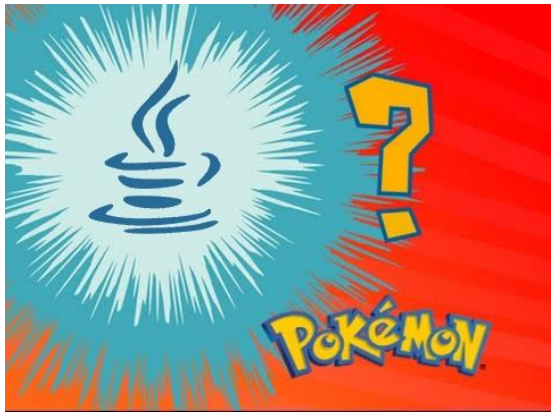
- Скорость работы
- Библиотека htsjdk

R

- Построение графиков
- Библиотека bioconductor

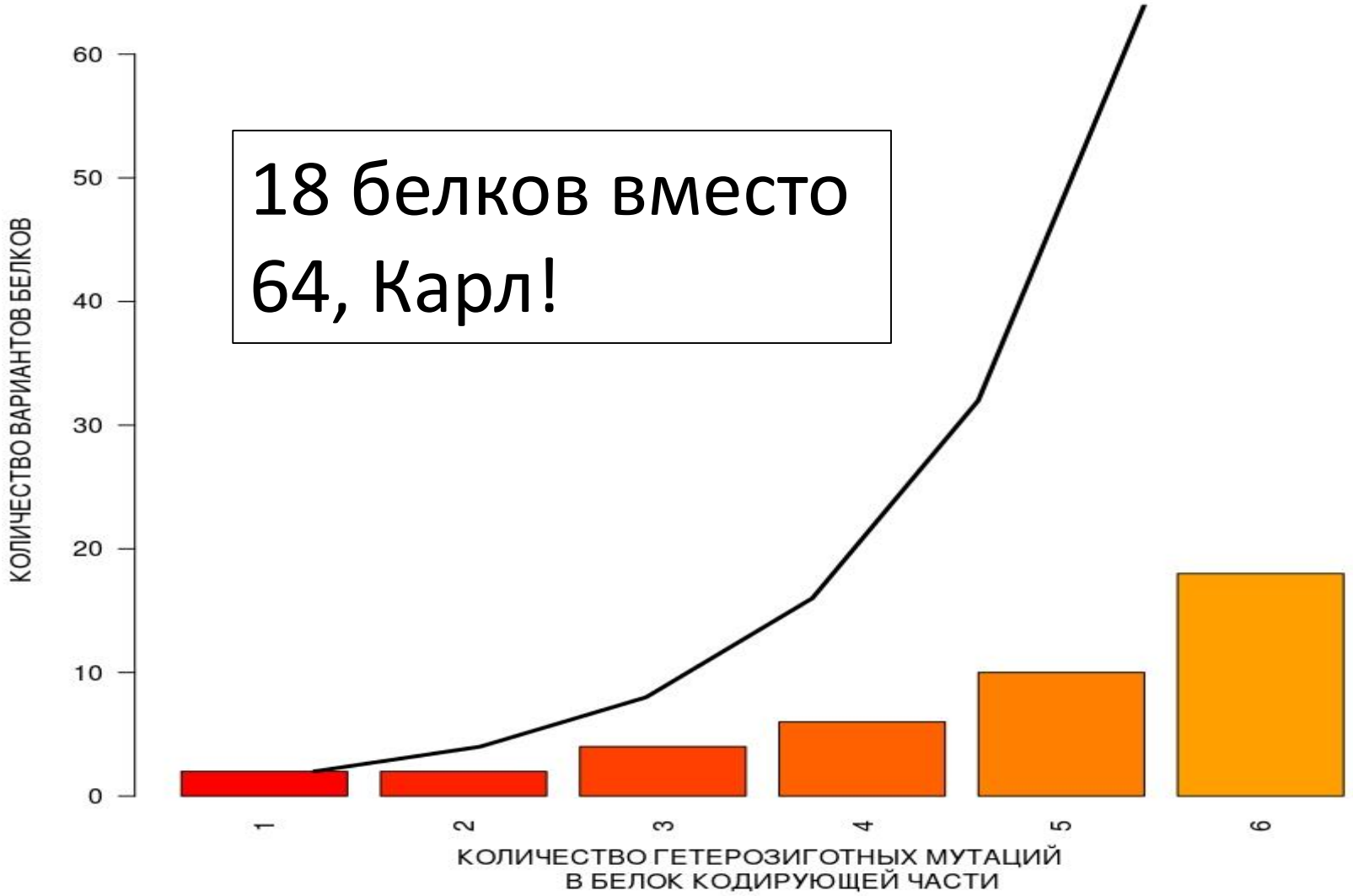
Библиотека ProRecon на языке Java

- Protein Reconstruction → ProRecon
- Полностью функциональна
- Выложена вместе с jar на github
- Покрыта тестами
- Качество существенно повышено за счёт сравнения результатов с полученными в python



https://github.com/Petr1990/protein_reconstruction

Количество гипотетических и возможных белков



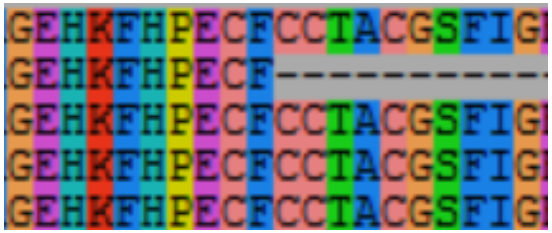
Произошедшие изменения белка

Делеция

Мутация перед стоп кодоном

Сдвиг рамки считывания

Мутация провоцирующая стоп кодон

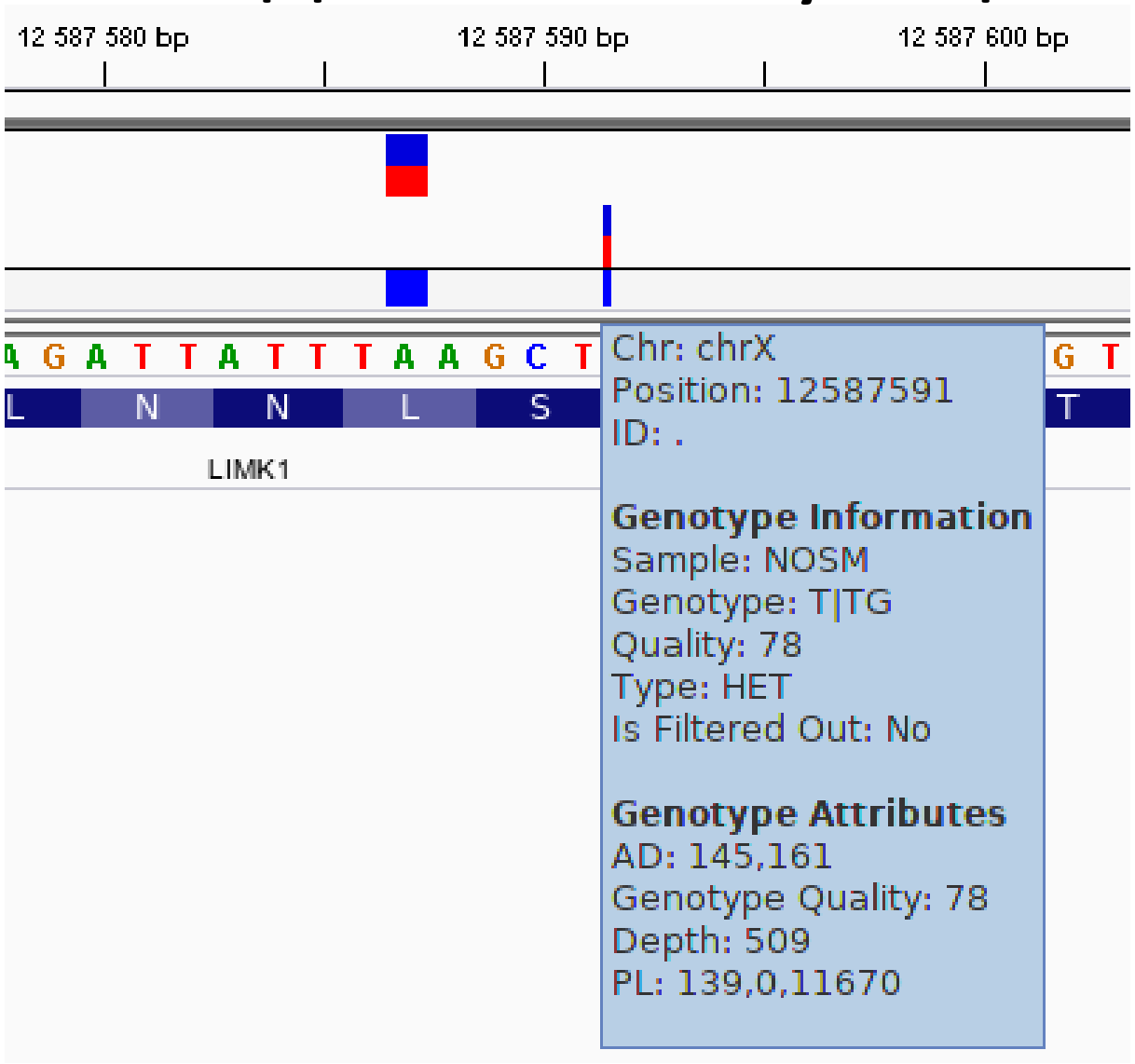


Планы на будущее

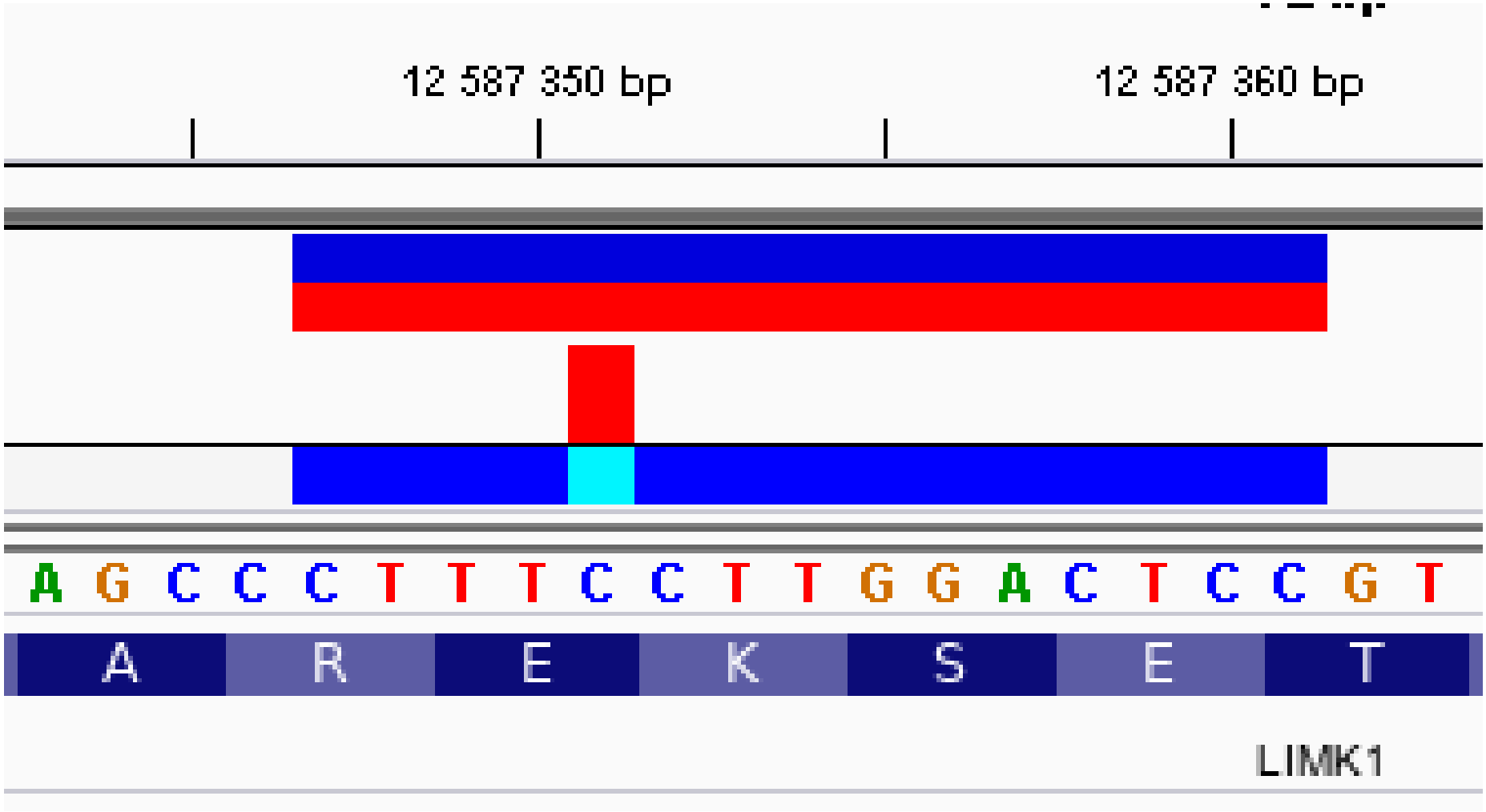
1) Учёт сложных случаев
   - Совпадение координат мутаций
   - Учёт сруктурных вариаций (инверсия, БНД)
   - Обработка мутаций затрагивающих сайты сплайсинга
2) Доработка качества библиотеки для полноценного релиза на гитхаб и хранилище Java библиотек
3) Интеграция в геномный браузер NGB
4) .....
5) PROFIT!!!

Спасибо за внимание!

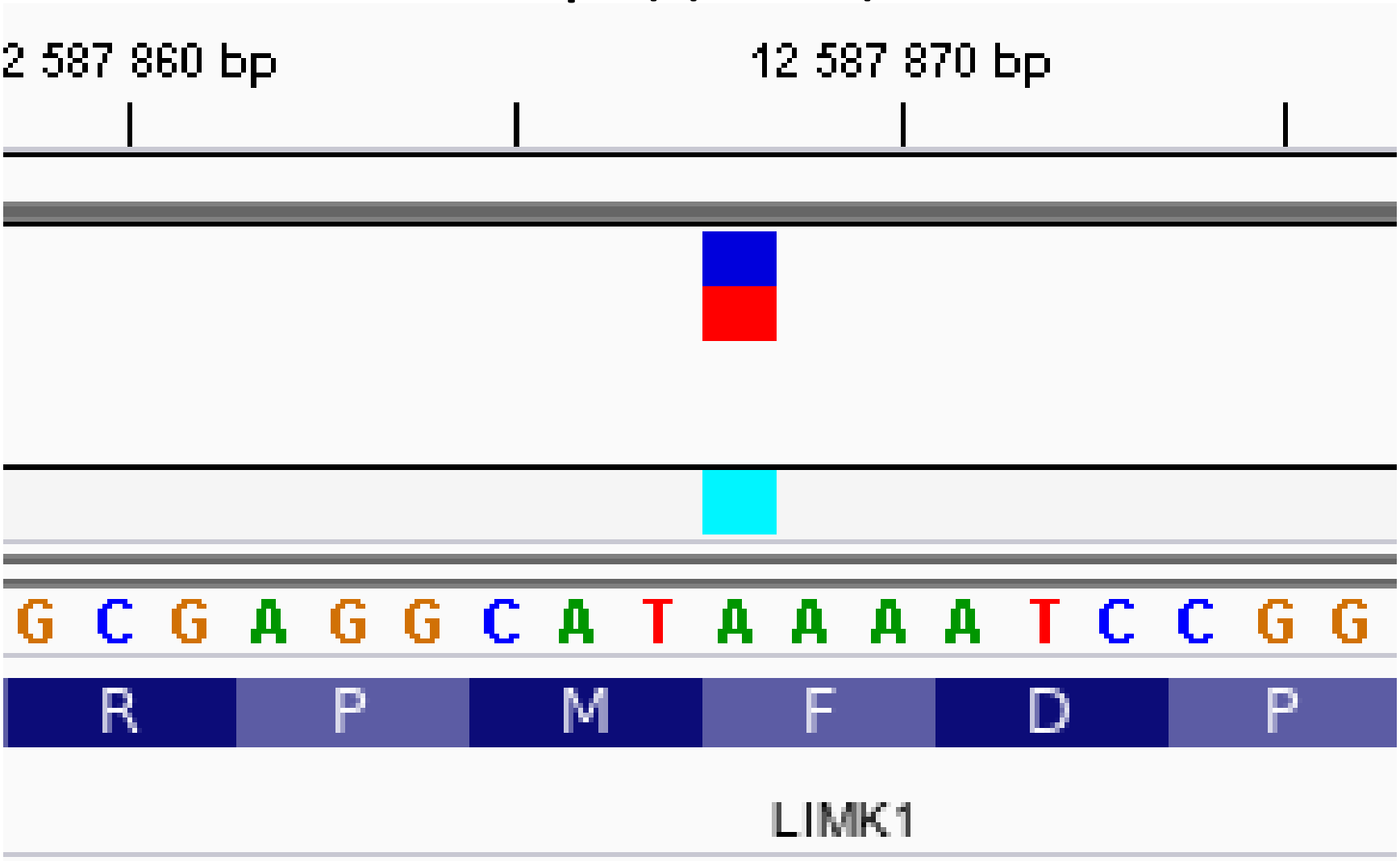
Взаимодействие мутаций



Наложение мутаций



Пример гетерозиготной фрейм-шифт делеции



Попытка выравнивания фрейм-шифта